Open access • Journal Article • DOI:10.1088/1741-2560/13/2/026019

Efficient mental workload estimation using task-independent EEG features

— Source link

Raphaëlle N. Roy, Sylvie Charbonnier, Sylvie Charbonnier, Aurélie Campagne ...+2 more authors

Institutions: University of Grenoble, Centre national de la recherche scientifique

Published on: 15 Feb 2016 - Journal of Neural Engineering (IOP Publishing)

Topics: Workload

Efficient mental workload estimation using task-independent EEG features

R N Roy[1,2], S Charbonnier[1,3], A Campagne[1,4] and S Bonnet[1,2]

[1] Univ. Grenoble Alpes, F-38000 Grenoble, France
[2] CEA, LETI, MINATEC Campus, F-38054 Grenoble, France
[3] CNRS, Gipsa-Lab, F-38000, Grenoble, France
[4] CNRS, LPNC, F-38000, Grenoble, France

E-mail: roy.raphaelle@gmail.com

Abstract
Objective. Mental workload is frequently estimated by EEG-based mental state monitoring systems. Usually, these systems use spectral markers and event-related potentials (ERPs). To our knowledge, no study has directly compared their performance for mental workload assessment, nor evaluated the stability in time of these markers and of the performance of the associated mental workload estimators. This study proposes a comparison of two processing chains, one based on the power in five frequency bands, and one based on ERPs, both including a spatial filtering step (respectively CSP and CCA), an FLDA classification and a 10-fold cross-validation. *Approach*. To get closer to a real life implementation, spectral markers were extracted from a short window (i.e. towards reactive systems) that did not include any motor activity and the analyzed ERPs were elicited by a task-independent probe that required a reflex-like answer (i.e. close to the ones required by dead man's vigilance devices). The data were acquired from 20 participants who performed a Sternberg memory task for 90 min (i.e. 2/6 digits to memorize) inside which a simple detection task was inserted. The results were compared both when the testing was performed at the beginning and end of the session. *Main results*. Both chains performed significantly better than random; however the one based on the spectral markers had a low performance (60%) and was not stable in time. Conversely, the ERP-based chain gave very high results (91%) and was stable in time. *Significance*. This study demonstrates that an efficient and stable in time workload estimation can be achieved using task-independent spatially filtered ERPs elicited in a minimally intrusive manner.

Keywords: mental workload, EEG, spatial filtering, classification, time-on-task

(Some figures may appear in colour only in the online journal)

1. Introduction

Mental state monitoring (MSM) through physiological computing, or neuroergonomics, is an actively growing research field, for it possesses numerous human factor applications, ranging from safety (e.g. car driving, nuclear plant monitoring), to smart technology development (Fairclough 2008, Parasuraman *et al* 2012). The new tools of neuroergonomics are passive Brain-Computer Interfaces (pBCIs). Those systems make use of an operator's neural activity in order to implicitly enhance human–machine interaction (George and Lécuyer 2010, Zander and Kothe 2011, van Erp *et al* 2012). Several mental states are currently the focus for research, including mental fatigue, attention, and affective states. Amongst them is mental workload, which is extensively documented in the literature, from its neural correlates to the processing chains that allow its estimation. Mental workload is defined quite differently depending on authors and study fields. Generally, it is considered that workload reflects task difficulty and the associated mental effort (Gevins and Smith 2007), or as Young *et al* (2015) state it in their review, mental workload is 'a multidimensional construct [...] determined by characteristics of the task (e.g. demands, performance), of the operator (e.g. skill, attention), and to a degree, the environmental context in which the performance occurs.' The task difficulty can be characterized in terms of quantity of engaged cognitive resources. This resource engagement to answer a given difficulty can correspond to i) an increase in short-term or working memory load for one specific task, ii) an increase in the number of items or tasks to process in parallel, or even iii) to the speed at which a task has to be performed (i.e. temporal pressure).

1.1. Workload cerebral markers

Amongst the methods that allow direct mental state assessment, electroencephalography (EEG) has become the dominant tool for BCIs (Nicolas-Alonso and Gomez-Gil 2012), and therefore for pBCIs, thanks to its low cost, high temporal resolution and ease of use. Several markers can be extracted from the EEG data, such as spectral markers and event-related potentials (ERPs).

Hence, workload can be estimated from electroencephalographic spontaneous activity using spectral markers. Indeed, it influences the EEG band power in several frequency bands. For instance, when load increases a decrease in alpha power (8–12 Hz) at centro-parietal sites has been reported, jointly with an increase in theta (4–8 Hz), and even delta power (1–4 Hz) at centro-frontal sites (Schober *et al* 1995, Gevins and Smith 2000, Missonnier *et al* 2006, Gomarus *et al* 2006, Holm *et al* 2009, Stipacek *et al* 2003, Antonenko *et al* 2010, Roy *et al* 2013). Several studies also reported variations with task difficulty and workload in the gamma band power (>30 Hz; Koles and Flor-Henry 1981, Dussault *et al* 2005, Berka *et al* 2007, Ossandón *et al* 2011).

Workload can also be estimated using neurophysiological markers temporally linked to stimulations, such as ERPs. Several early and late ERP components have been reported to be sensitive to workload modulations. As attested by various studies that used concurrent target detection tasks, the amplitude of the P300 component evoked by targets of concurrent tasks decreases with increasing workload (Natani and Gomer 1981, Kok 2001, Schultheis and Jameson 2004, Gomarus *et al* 2006, Holm *et al* 2009) and is thought to be a reliable indicator of working memory load resource allocation (Kok 2001, Fu and Parasuraman 2007). The amplitude of the early N1, N2 and P2 components is also reduced when workload increases (Ullsperger *et al* 2001, Allison and Polich 2008, Miller *et al* 2011). What's more, regarding the ERPs of items to recall, all their components from 150 ms to 800 ms are attenuated at the Cz electrode when workload increases (Gomarus *et al* 2006).

1.2. Workload estimation

Mental workload estimation can be performed thanks to tools that have been developed for active BCIs. Thus, most of the processing chains dedicated to workload estimation that are reported in the literature include a feature extraction step (e.g. frequency filtering or event-related potentials) and a translation step (e.g. classification). Additionally, spatial filtering techniques commonly used for active BCI applications have recently been applied to enhance mental workload estimation. The spatial filtering step enables a reduction of the dimensionality of the feature space, as well as an enhancement of the discriminability between classes.

Regarding spectral markers, the research literature reports the use of a principal component analysis (PCA) or a common spatial patterns (CSP) filtering step. The PCA is a method that finds uncorrelated components without taking the label of the data into account (Pearson 1901). On the other hand, the CSP method determines spatial filters that maximize the variance of the spatially filtered signals of one class while minimizing it for the other class (Blankertz *et al* 2008). Several authors used a CSP filtering to efficiently enhance the performance of their online workload estimation, such as George *et al* (2012) or Schultze-Kraft *et al* (2013), yet they used long analysis or decision windows. Regarding systems that use shorter analysis windows with spectral markers, Mühl *et al* (2014) obtained 73% of correct binary classifications with 2 s windows using a CSP filtering. Lastly, to our knowledge, the best performances using the shortest window length (1 s) were obtained by Heger *et al* (2010)–with a PCA filtering- and Dijksterhuis *et al* (2013) -with a CSP filtering- with binary classification accuracies above 80%.

However, it should be noted that most studies that report very high classification accuracies based on spatially filtered spectral markers actually discriminate between either a relaxed state vs. an engaged state (e.g. Heger *et al* 2010), or two engaged states but one that involves very little motor activity and the other one important motor activity (e.g. Schultze-Kraft *et al* 2012, Dijksterhuis *et al* 2013). Therefore, in the first case it would be more accurate to say that the processing chain estimates task engagement rather than mental workload. For even though mental workload can be defined for some authors as an engagement of cognitive resources to perform a given task, to us the reverse isn't necessarily true. The fact of engaging cognitive resources (e.g. when engaging in a task) doesn't allow us to say that one has a higher workload than in its previous state. It might be true, but we do not know it for sure. It would be better to evaluate mental workload levels between two conditions in which only one process is modulated (e.g. working memory load). In the second case, in our view the estimation is potentially biased by the presence of this motor activity. Therefore, it seems that the high performances that were achieved may have mainly relied on the detection of this motor activity.

Until even more recently than for spectral markers, spatial filtering had never been applied to ERPs for mental workload assessment. Yet it proved very efficient. Indeed, for a binary classification between two 'real' workload levels Brouwer *et al* (2012) obtained only 64% of accuracy with the raw signal from 7 electrodes, whereas Mühl *et al* (2014) applied a Fisher Spatial Filtering (FSF; Hoffmann *et al* 2006) method to the non-target ERPs of their n-back memory task and reached 72% of correct binary classifications. Moreover, Roy *et al* (2015a, 2015b) demonstrated that a binary 'real' workload estimation based on the ERPs of the test items of a Sternberg memory task was significantly enhanced when using an xDAWN (Rivet *et al* 2009) or a canonical correlation analysis (CCA; Hotelling 1936) filtering step with performances above 95% of correct binary classifications.

1.3. Real-life applicability

Spectral markers have a great advantage against ERPs for they enable continuous and non-intrusive monitoring. Indeed, the elicitation of ERPs requires the use of stimulations, or probes, which can be quite distracting. Also, the literature on effective mental estimation makes only use of task-dependent probes. However, for real life applications, ERP-based systems should only use task-independent probes. What's more, in order to reduce the cognitive resources required to perform this task, and to avoid a double-task situation, the answer given by the participant to the appearance of such probes should be as reflex-like as possible, similar to the reflex answers from train and tramway conductors awaited by dead man's vigilance control devices. To our knowledge no literature exists regarding the use of such minimally intrusive task-independent probes for mental workload classification.

Secondly, in order to progress towards mental state monitoring systems that can be implemented in real life settings, several major problems remain to be solved. As it has recently been stressed out, amongst those problems is the impact of other mental states, such as stress or mental fatigue (which arises for instance from growing time-on-task for tasks that require sustained attention; Boksem *et al* 2005), that can induce major non-stationarity in the EEG signal (Mühl *et al* 2014, van Erp *et al* 2012). Mühl *et al* (2014) have shown that a social stress present either during the training or the testing sessions significantly degrades classification performances, with a stronger impact when the processing chain is based on ERPs than based on spectral markers. Moreover, Roy *et al* (2013) have demonstrated that spectral markers were subject to time-on-task induced non-stationarity. Yet there is still a lack of literature regarding the impact of time-on-task and mental fatigue on ERPs and their associated processing chain' performance, as well as on the impact of other mental states on workload estimation.

1.4. Current study

To our knowledge, no study has ever compared the workload estimation performance of chains based on spectral markers or event-related potentials from task-independent probes within the same experimental protocol. Moreover, no study has compared these performances when time-on-task—and mental fatigue- increases. This study was designed to meet the lack of literature on this issue. The contributions of this paper are two-fold:

i. It performs a comparison of two well employed EEG types of markers, namely spectral markers and event-related potentials, for a *real* mental workload estimation (i.e. between two levels of workload and not between relaxed vs. engaged states) using enhanced processing chains that include a spatial filtering step. The time windows used to extract each marker depend on its type. They are selected for each marker in order to obtain the best performance. Spectral markers are extracted without any motor activity bias, and ERPs are elicited using minimally intrusive task-independent probes;
ii. It assesses the proneness to time-on-task induced non-stationarity of these two types of markers, both at the neurophysiological level and at the processing chain performance level.

2. Materials

2.1. Ethic statement

This research was promoted by Grenoble's clinical research direction (France) and was approved by the French ethics committee (ID number: 2012-A00826-37) and the French health safety agency (B120921-30). It was conducted according to the principles expressed in the Helsinki Declaration.

2.2. Participants

Twenty healthy volunteers (25 years old $\pm$ 3.5 years; 9 females) performed the experiment. They were right handed, had normal or corrected-to-normal vision, had no neurological or psychiatric disorders, nor were they under any medication. They signed a written consent and received an 80-euro compensation.

2.3. Stimuli

The stimuli were displayed against a grey background using the E-prime software (E-prime Psychology Software Tools Inc., Pittsburgh, USA) onto a 21-in. monitor (with a 1024×768 pixels resolution and a 75 Hz refresh rate) located 70 cm from the participants (approximately 2° of visual angle). They consisted of centered black digits (from 1 to 9) flanked with question marks, and filled geometrical shapes (triangle, circle, square and rhombus).

2.4. Experimental protocol

In this experimental protocol, two factors were manipulated: working memory load and time-on-task. First, working memory load was manipulated using a modified Sternberg task (Sternberg 1966). The participants had to memorize, for each trial, a different list of digits, presented sequentially and visually on the computer screen. After a 2 s time lapse, a test item flanked with question marks was presented for which the participants had to perform a recognition task (figure 1). Specifically, they had to indicate as quickly and as accurately as possible using a response box (two keys, both operated with the right hand) whether the probe had been presented or not in the list to memorize (50% of cases). Two levels of workload were considered, i.e. they had 2 or 6 digits to memorize, respectively corresponding to a low and a high

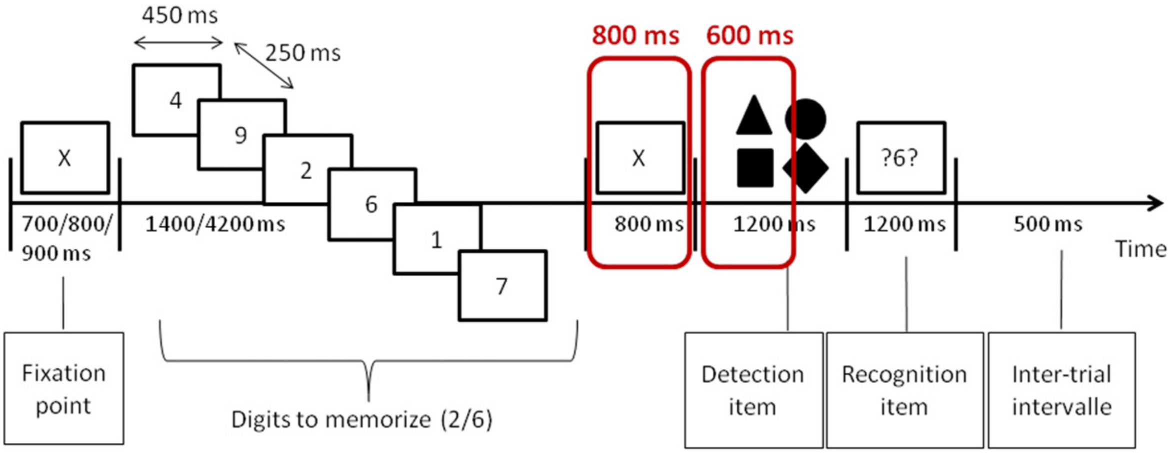

Figure 1. Trial structure. The circled segments were used for respectively the spectral markers and the event-related potentials' extraction and classification.

workload level. To assess the feasibility of estimating the workload generated by a given task using task-irrelevant probes, a detection task was inserted between the memorization and the recognition steps of the Sternberg task (figure 1). In this task, the participants had to detect the appearance of a geometrical shape (triangle, rhombus, circle or square) using their left hand. The geometrical shape was always presented and the same overt motor response was to be given regardless of the nature of the geometrical shape, therefore creating an automatic answer. Participants had to indicate as quickly and as accurately as possible that they had perceived the geometrical shape. This task was performed without interruption during two 9.7 min blocks, which each included 80 trials, half of which induced a low workload, the other half, a high workload. Trials of low and high workload were pseudo-randomly presented.

The second factor, time-on-task (TOT), was manipulated by the experimental paradigm during one long session. The two blocks that were analyzed were separated by at least 40 min, during which the participants performed another version of the task, more tiring, in which the detection task involved selective attention processes (i.e. selective detection of the triangle). Two levels of time-on-task (fatigue levels) were thus defined—a short TOT for the first block and a long TOT for the second block. Before completing the tasks, the participants performed two 5 min blocks of training, one for each task, in order to minimize learning effects in the study.

3. Methods

3.1. Data acquisition and preprocessing

Data acquisition was performed at IRMaGe Neurophysiology facility (Grenoble, France). For both experimental blocks, the participants' reaction times (RTs) and accuracy to the detection and recognition tasks were measured. In addition, participants' mental fatigue elicited by TOT was measured before, during (in the middle of the experience) and at the end of the experiment using the Karolinska Sleepiness Scale (KSS; Kaida *et al* 2006). Data points exceeding 2.5 standard deviations from the mean for each condition were excluded.

Moreover, participants' electroencephalographic (EEG) activity was continuously recorded using an Acticap® (Brain Products, Inc.) equipped with 32 Ag-AgCl unipolar active electrodes that were positioned according to the 10–20 system. The reference and ground electrodes used for acquisition were those of the Acticap, i.e. FCz for the reference, and AFz for the ground. The electro-oculographic (EOG) activity was also recorded using 2 electrodes positioned at the eyes outer canthi, and 2 respectively above and below the left eye. Participants were instructed to limit blinking and eye-movements both during the two fixation crosses and item presentation. Impedance was kept below 10 kΩ for all electrodes. The signal was amplified using a BrainAmp™ system (Brain Products, Inc.) and sampled at 500 Hz with a 0.1 Hz high-pass filter and a 0.1 μV resolution. The digital EEG signal was band-pass filtered between 1 and 40 Hz, and re-referenced to a common average reference. Artifacts related to ocular movements (saccades and blinks) were corrected using the signal recorded from the EOG electrodes and the SOBI algorithm (Belouchrani *et al* 1997). All trials were kept for analysis.

3.2. Processing chains

Both processing chains for each type of marker, i.e. spectral markers and event-related potentials, included a segmentation step, a spatial filtering step, and a classification step using a Fisher Linear Discriminant Analysis (FLDA). Thus, in a general manner, the signal $\mathbf{X}$ (N_e—number of channels- x N_s—number of samples) of a given epoch was processed and filtered, resulting in the following signal $\mathbf{Z}$ (N_f—number of filters- x N_s): $\mathbf{Z} = \mathbf{W}^T\mathbf{X}$. $\mathbf{W}$ corresponds to the matrix that contains the spatial filters (column vectors; N_e x N_f). The vectors of the matrix $\mathbf{A} = (\mathbf{W}^{-1})^T$ give the spatial patterns. The feature extraction step was performed on this signal $\mathbf{Z}$, giving as a result a feature vector $\mathbf{f}$. Details are given in the following sub-sections regarding the analysis windows, spatial filtering steps and feature extraction procedures for each type of marker.

3.2.1. Spectral markers. In order to compute the spectral markers to use for classification, the 800 ms segment corresponding to the second fixation cross was extracted from each trial (**X**). This segment was chosen because it contains no motor activity what so ever, and should mostly reflect memory retention processes. It is very short in order to get closer to a real-life implementation in which a reactive system would be needed. Next, the data were filtered using a 5-order Butterworth filter in each of the 5 following frequency bands: delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (13–30 Hz) and gamma (>30 Hz). The number of electrodes was then reduced by keeping only the $k = 15$ electrodes that allowed the Riemannian distance between the 2 classwise-mean covariance matrices to be maximal. Introducing the $k \times k$ principal submatrix $\Sigma^{(k)}$, this technique amounts to estimate the set of k indices that maximize the Riemannian distance $\delta_R \Sigma_1^{(k)}, \Sigma_2^{(k)}$. For more details on the method see Barachant and Bonnet (2011).

Then, the data **X** were spatially filtered using 6 CSP filters. The 3 pairs of filters with the highest positive and negative eigenvalues were kept. They were computed by maximizing the ratio of the mean covariance matrices of each class:

$$\mathbf{w}_{\mathrm{CSP}} = \mathrm{argmax}_{\mathbf{w}} \frac{\mathbf{w}^T \Sigma_1 \mathbf{w}}{\mathbf{w}^T \Sigma_2 \mathbf{w}} \; s.\, t. \;\; \mathbf{w}^T (\Sigma_1 + \Sigma_2)\, \mathbf{w} = 1$$

This corresponds to a Rayleigh quotient and the potential solutions which comply with the following equation: $\boldsymbol{\Sigma}_1 \mathbf{w}_{\mathrm{CSP}} = \lambda \boldsymbol{\Sigma}_2 \mathbf{w}_{\mathrm{CSP}}$. This problem can be solved by generalized eigenvalue decomposition. The electrode selection and spatial filter computation steps were performed using the training set and were then directly applied on the testing set.

Lastly, the features were extracted from the filtered signal **Z** by taking the log variance of the filtered signals for each of the 5 bands. Therefore, the feature vector **f** for one trial had a 30×1 dimension (i.e. 5 bands $\times$ 6 values).

For the group level analysis of markers, the mean power in each band, as well as the low alpha (8–10 Hz) and high alpha (10–12 Hz) power, were extracted on the segment **X** for all midline electrode sites using Welch's power spectral density estimation. We chose to present the results at the neurophysiological level for these two alpha sub-bands because there are some references that point towards a major modulation due to time-on-task of the low alpha sub-band power and not particularly of the high alpha sub-band (Gale *et al* 1977).

3.2.2. Event-related potentials. In order to extract the event-related potentials to use for classification, the 600 ms segment corresponding to the task-independent probe that is the detection item was selected. The data was decimated to 100 Hz using a 5 point moving average. Next, the data was baseline corrected by subtracting to it the mean amplitude of the preceding 200 ms.

Then, the data **X** were spatially filtered using CCA filters. Only the two spatial filters with the highest associated eigenvalue were selected. As Spüler *et al* (2014) detailed it, in a two-class scenario the CCA filters are computed in order to maximize the correlation between the EEG signal **X** and the matrix that contains the average ERP response for each class successively replicated for each temporal segment respective to its class label and then concatenated. Therefore, this problem can also be seen as maximizing the ratio of the covariance matrices of **X** and **Z**, the filtered signal:

$$\mathbf{w}_{\mathrm{CCA}} = \mathrm{argmax}_{\mathbf{w}} \frac{\mathbf{w}^T \Sigma_{\mathbf{Z}} \mathbf{w}}{\mathbf{w}^T \Sigma_{\mathbf{X}} \mathbf{w}}$$

The solutions to this Rayleigh quotient can therefore be computed in the same way as for the CSP filtering.

Finally, the features were extracted from **Z** by keeping the filtered signals that can be referred to as virtual electrodes. Therefore, the feature vector **f** for one trial had a 120×1 dimension (i.e. 60 samples $\times$ 2 virtual electrodes).

For the group-level analysis of markers, the average peak amplitude of the N1, P2, N2 and P3 ERP components was extracted at all midline electrode sites.

3.3. Validation criterion & analyses

The performance of each processing chain was assessed based on its intra-subject binary classification accuracy with a 10-fold cross validation procedure. For the first analysis, the training and testing sets were both extracted from the first block (short TOT), which produced 40 epochs of low mental workload and 40 of high mental workload. For the second analysis, the training set was extracted from the first block and the testing set from the second block (long TOT), also formed of 40 epochs of low mental workload and 40 of high mental workload. All spatial filters were learned on the training sets.

The impact of workload and time-on-task on the behavioral, subjective, physiological data, as well as on the performance data obtained from the two processing chains was statistically assessed. For every obtained performance, its difference from random was assessed using single-sample t-tests. Next, using repeated measures ANOVAs and Tukey post-hoc tests that correct for multiple comparisons, all performances were compared between themselves (i.e. one ANOVA: 2 marker types*2 workload levels*2 TOT levels). In the same way, the impact of the workload and time-on-task levels on the behavioral (response times and accuracy to the Sternberg task), subjective and physiological data was assessed (e.g. one ANOVA for spectral markers: 6 midline electrodes*7 bands*2 workload levels*2 TOT levels). The significance level was set at 0.05.

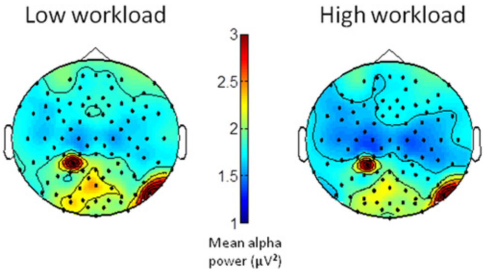

Figure 2. Topographical representation of the mean alpha power on the 800 ms window depending on workload condition (average across participants).

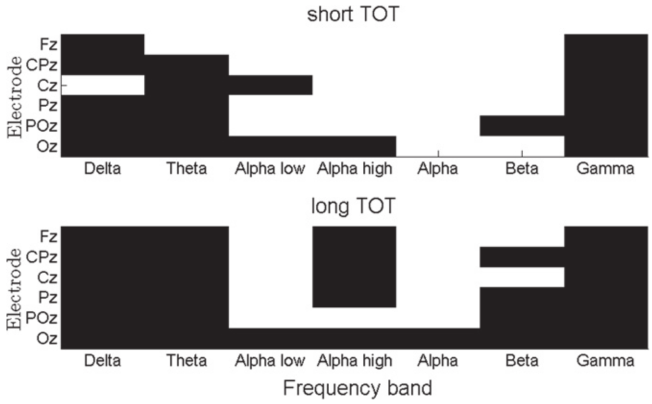

Figure 3. Impact of time-on-task (TOT) on the presence of significant differences in average power between workload levels per electrode and per band at the group level. White: significant difference ($p < 0.05$).

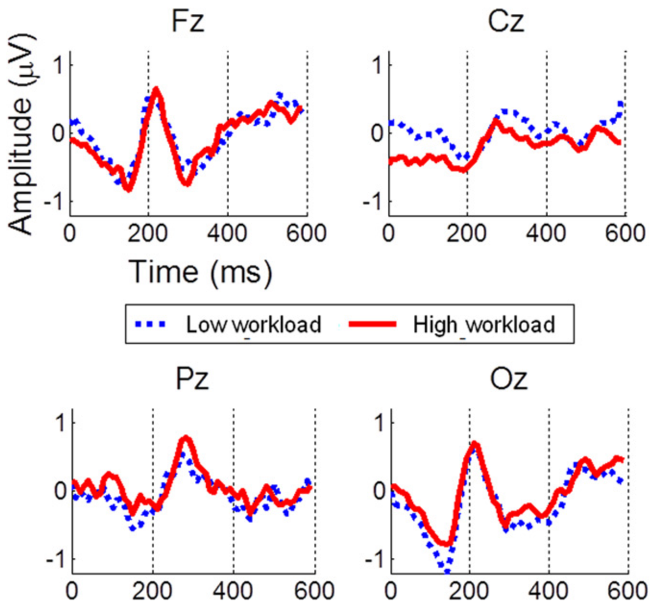

Figure 4. Grand average event-related potentials of the task-independent probe at major midline electrode sites.

4. Results

The results obtained using the previously described statistical analyses are detailed in the following subsections. First, the behavioral results are described, in order to assess the efficiency of the experimental protocol to manipulate both workload and time-on-task effects. Next, to determine if these factors do impact classical neurophysiological markers, the results obtained on spectral markers and event-related potentials are detailed. Then, the spatial patterns of each processing chain are described as an indication of the topographical origin of the enhanced EEG activity for mental workload classification. Lastly, to evaluate whether one of the markers allows a better estimation of workload level, the estimation results of each chain are compared. Then, the impact of time-on-task on these chains is assessed as an additional indication of the efficiency of each marker.

4.1. Behavioral and subjective data

For the recognition task, the statistical analysis uncovered a main effect of workload on both reaction times ($F(1,19) = 57.02$, $p < 0.001$) and accuracy ($F(1,19) = 14.63$, $p < 0.01$). With increasing workload, the participants were longer to recognize the test item ($m1_{RT} = 514.66$ ms; $sd1_{RT} = 70.14$ ms; $m2_{RT} = 593.70$ ms; $sd2_{RT} = 67.66$ ms) and were also less accurate ($m1_{ACC} = 0.95$; $sd1_{ACC} = 0.05$; $m2_{ACC} = 0.91$; $sd2_{ACC} = 0.07$). No other statistical effect was observed for this task.

For the detection task, participants were slower to answer when time-on-task increased ($F(1,19) = 8.84$, $p < 0.01$; $m1_{RT} = 378.5$ ms; $sd1_{RT} = 65.91$ ms; $m2_{RT} = 387.2$ ms; $sd2_{RT} = 73.12$ ms). No other statistical effect was observed for this task.

Moreover, the participants' mental fatigue elicited by TOT was measured before (1), during (2) and after the experiment (3). They reported feeling increasingly tired as the experiment progressed in time ($F(2,38) = 50.06$, $p < 0.01$; $m1_{KSS} = 3.3$; $sd1_{KSS} = 0.8$; $m2_{KSS} = 5.1$; $sd2_{KSS} = 1.62$; $m3_{KSS} = 6.25$; $sd3_{KSS} = 1.65$).

4.2. Electrophysiological data

4.2.1. Spectral markers. At the group level, the statistical analyses revealed a significant decrease in alpha and beta power at all midline electrode sites when workload increased ($p < 0.05$; n.s. for the other frequency bands). This phenomenon is illustrated by figure 2 that displays the topographical distribution of the mean absolute power in the alpha band depending on the workload level condition averaged across participants. It can be seen that the alpha power indeed decreases when workload increases in a more pronounced manner at central electrode sites.

As regards the impact of time-on-task on these spectral markers at the group level, there was a reduction in the number of significant differences in the mean power between workload levels when time-on-task increased. Indeed, at most electrode sites several bands such as the alpha high and the

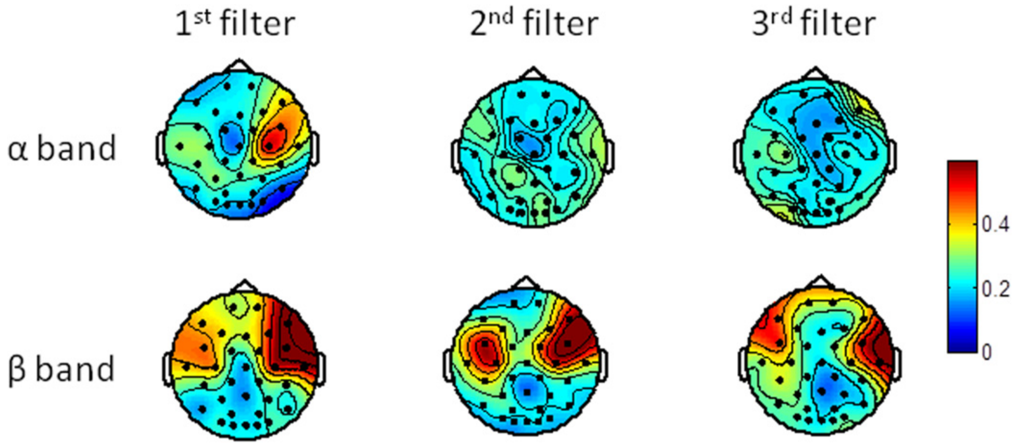

Figure 5. First 3 CSP spatial patterns for the alpha and beta bands, extracted from the 800 ms window (average across participants).

beta bands lose their discriminative power at the group level when TOT increases (figure 3).

4.2.2. Event-related potentials. Figure 4 gives the grand-average ERPs across participants for the task-independent probe at major midline electrode sites. The typical components reported to be modulated by workload can be noticed, i.e. N1, P2, N2 and P3 (Gomarus *et al* 2006, Allison and Polich 2008, Holm *et al* 2009). However, the statistical analyses revealed no significant effect of workload on these ERP components' amplitude at the group level ($p > 0.05$). In the same way that these raw markers do not reflect any workload impact at the group level, TOT had no impact on them at this level.

4.3. Spatial patterns

Both processing chains used for mental workload estimation included a spatial filtering step. Figure 5 displays the averaged first 3 CSP spatial patterns for the alpha and beta bands. The spatial patterns inform us about the source of the enhanced activity. It can be seen that the activity that was enhanced by these filters mainly originates from lateral fronto-central sites. As for the CCA filtering, figure 6 displays the two averaged CCA spatial patterns that were used for the ERPs elicited by the task-independent probe. It can be seen that the activity that was enhanced by these filters originate from central and occipital regions.

4.4. Classification results

Regarding the workload estimation achieved thanks to the two types of markers and their associated processing chain, generally the chain based on the ERPs gave significantly better results than the chain based on the spectral markers ($p < 0.001$). Indeed, as reported by figure 7, the use of the spectral markers and their associated processing chain gave rise to low estimation performances, with only 60% of mean accuracy for the 20 subjects. On the other hand, the use of the ERPs elicited by the task-independent probe and their

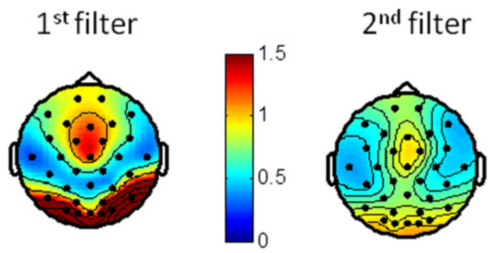

Figure 6. CCA spatial patterns for the event-related potentials of the task-independent probe (average across participants).

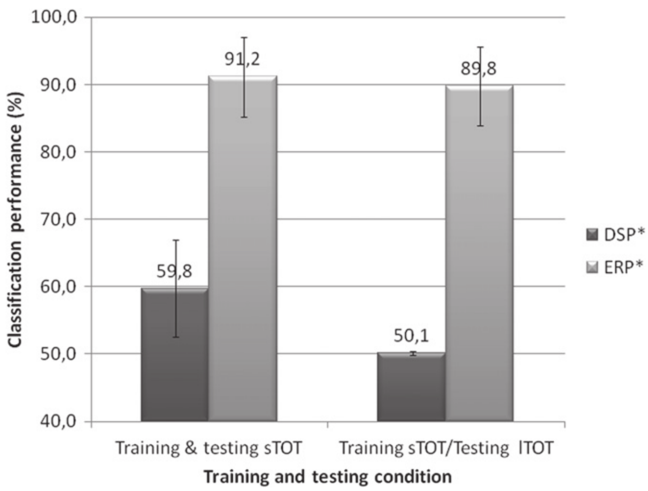

Figure 7. Classification results depending on the training and testing condition (average and standard deviation across participants). *DSP*: processing chain based on the spectral power markers; ERP*: processing chain based on the event-related potentials from the task-independent probe; sTOT: short time-on-task; lTOT: long time-on-task.*

associated processing chain gave rise to very high performances, with as much as 91% of mean accuracy.

Furthermore, the performance achieved using the chain based on the spectral markers significantly dropped when the chain was trained on trials from the beginning of the session and tested on trials at the end of the session ($p < 0.001$). However, this impact of time-on-task was not present for the performance achieved using the chain based on the ERPs ($p = 0.61$). Lastly, all obtained performances were significantly higher than random ($p < 0.001$), except the one obtained with the chain based on the spectral markers and tested on trials from the end of the session ($p = 0.30$).

The negative impact of TOT on the workload estimation performances obtained using the chain based on the spectral markers can be explained by the fact that these markers are already impacted by TOT at the neurophysiological level. On the other hand, the performances obtained using the chain based on the ERPs are not afflicted by this impact, in the same way that the ERPs were not impacted by TOT at the neurophysiological level. In any case, these results reveal that the chain based on the ERPs is both more efficient in a general manner, but is also stable in time whereas the chain based on spectral markers is not.

5. Discussion

This study had one major purpose: to directly compare the relevance of two classical EEG markers for efficient and minimally intrusive mental workload estimation, a classical baseline measure in the spectral domain and the event-related potentials elicited by probes. The experimental protocol that was used to manipulate workload allowed a *real* difference in workload to appear between conditions. Indeed, in the low workload condition participants had to memorize 2 digits whereas in the high workload condition they had to memorize 6 digits. The EEG markers that were investigated were spectral markers –i.e. the power in the 5 classical bands delta, theta, alpha, beta and gamma- and event-related potentials –i.e. the EEG activity elicited by a given stimulation and time-locked to its appearance. In order to get closer to a real life implementation, the spectral markers were extracted from a very short analysis window (800 ms) that would allow monitoring systems to be quite reactive. Most importantly, the selected analysis window did not include any motor activity in order to avoid biases in the analyses and can therefore be considered a good baseline measure. Moreover, the event-related potentials were elicited by task-independent probes that required a reflex-like response from the participants, akin to the one performed by train conductors in response to warnings implemented in dead man's vigilance control devices. This way, the probes were minimally intrusive.

First, this study presented the impact of the workload level on these markers, and also provided a direct comparison of the relevance of these two types of markers for workload estimation. This estimation was performed thanks to processing chains that included a spatial filtering step. At the group level, there was a significant modulation of the average power in the alpha and beta frequency bands due to workload in accordance with the literature (Schober *et al* 1995, Gevins and Smith 2000, Stipacek *et al* 2003, Missonnier *et al* 2006, Gomarus *et al* 2006, Holm *et al* 2009, Antonenko *et al* 2010). In contrast, at the group level no effect was found for the event-related potentials, although modulations of components' amplitude were expected given the literature (Allison and Polich 2008, Ullsperger *et al* 2001, Boonstra *et al* 2013). This could be due to the fact that these studies report significant modulations on event-related potentials that are elicited by task-*dependent* probes. Moreover, even though there was no impact at the markers' level for the group of participants, there was an effective modulation of these markers per participant, since the intra-subject classifiers were efficient.

Regarding their efficiency for workload estimation, the chain based on the spectral markers gave only 60% of correct classifications. This is largely inferior to the results reported by Heger *et al* (2010). These researchers report classification accuracies above 80% using a 1 s analysis window. However, they did not perform a *real* workload estimation, as they only estimated whether the participants were engaged in a task or were relaxed. Therefore, our results had better be compared with the ones reported by Brouwer *et al* (2012) or by Grimes *et al* (2008), who both obtained 65% of correct classifications for a *real* binary workload estimation using an analysis window of more than twice our own in length (respectively 2,5 and 2 s). Therefore, our results using spectral markers allow a quicker estimate of workload with performances in line with the literature. Yet the performances obtained with this type of marker remain low. It should be noted that if, for a given system, reactivity is not a major criterion, then a longer window should be preferred as it allows a better estimation of the power in low frequency bands. Conversely, the ERP-based chain gave very good results with as high as 91% of correct classifications, which was significantly higher than the results obtained using the spectral markers. This is also largely superior to the performances reported by the literature using raw or spatially filtered ERPs elicited by task-*dependent* probes (Brouwer *et al* 2012, Mühl *et al* 2014). Therefore, it seems that our chain that includes a CCA filtering enables very efficient workload estimation even when using task-*independent* probes.

Hence, significant differences in power values were found at the group level while the classification results obtained using these features at the subject level were quite low. The opposite pattern was found for the probe item ERPs with no significant difference at the group level for the ERP components' amplitude, but high classification results at the subject level. This may be due to more between-subject variance for the ERPs than for the spectral markers that may stem from differences in motor response. An interesting subsequent analysis would be to perform a spectral analysis on the task-independent probe window, as well as an ERP analysis on the retention period. Moreover, the use of probes that do not require any overt response and of spectral markers on larger windows that do encompass motor activity should be thoroughly investigated. This way, any influence of motor activity would be explicitly analyzed.

Secondly, this study assessed the proneness to time-on-task induced non-stationarity of these two types of markers, both at the marker level and at the processing chain performance level. To our knowledge, our study is the first one to tackle this issue. At the marker level, there was a significant impact of time-on-task on the differences in average power from several frequency bands between workload levels.

Table 1. Advantages and limitations of each EEG marker type.

	Spectral markers	Event-related potentials
Advantages	Non intrusive	High performance
	Continuous monitoring	Stability in time
Limitations	Low performance	Intrusive
	Not stable in time	Discontinuous monitoring

However, the ERPs of the task-independent probes were unaffected. In the same manner, at the processing chain's level, the chain based on the spectral markers gave results that dropped to chance level once it was tested on trials from the end of the session, whereas the ERP-based chain was stable. Therefore, it appears here that spectral markers and their associated processing chain are less stable in time than ERPs.

It stems from this study that event-related potentials appear to be more efficient for mental workload estimation in a close to real life implementation than spectral markers, given that they provide better classification accuracies and are stable in time both at the marker level and at the estimation level. However, the use of these markers requires an external stimulation that can be disturbing for the participants. It is therefore less practical and more intrusive than spectral markers. In addition, spectral markers enable continuous monitoring whereas ERPs do not. The advantages and limitations of each marker type that result from this study are listed in table 1. In order to render the use of the ERPs less intrusive, a way would be to use *infrequent* task-independent probes that are *ignored* by the participants, i.e. do not require any overt or covert response. This idea was presented by Allison and Polich (2008) who assessed the impact of workload on the ERPs elicited by counted or ignored auditory probes using a single stimulus paradigm. Although they showed that workload could impact these ERPs at the group level, they did not pursue this study at the estimation level.

6. Conclusion

This study presented a direct comparison between two classical EEG marker types -spectral markers and event-related potentials- for mental workload estimation in a close to real life implementation. It appeared that event-related potentials elicited by task-independent probes and their associated processing chain that included a CCA filtering gave the best results with classification accuracies of 90%, which were stable in time. In comparison, spectral markers were prone to a time-on-task induced non-stationarity which rendered them useless for workload estimation, as the classification accuracies of their associated processing chain dropped from 60% to chance level when tested on trials from the end of the session. Therefore, despite the fact that it only allows for a discontinuous monitoring, it appears that the use of ERPs enables a more efficient estimation of workload than spectral markers in a close to real life implementation. Moreover, as we presented, its intrusiveness can be reduced by using minimally intrusive task-independent probes that only require a reflex-like response akin to that performed by train conductors in response to warnings from dead man's vigilance control devices. This work should be pursued by assessing the usability of *ignored* task-independent probes for mental workload estimation.

Acknowledgments

Grenoble Neurophysiology facility IRMaGe was partly funded by the French program 'Investissement d'Avenir' run by the 'Agence Nationale pour la Recherche': Grant 'Infrastructure d'Avenir en Biologie Santé' (ANR-11-INBS-0006).

References

Allison B Z and Polich J 2008 Workload assessment of computer gaming using a single-stimulus event-related potential paradigm *Biol. Psychol.* **77** 277–83

Antonenko P, Paas F, Grabner R and Gog T 2010 Using electroencephalography to measure cognitive load *Educational Psychology Review* **22** 425–38

Barachant A and Bonnet S 2011 Channel selection procedure using Riemannian distance for BCI applications *Proc. IEEE Conf. Neural. Eng.* pp 348–51

Belouchrani A, Abed-Meraim K, Cardoso J-F and Moulines E 1997 A blind source separation technique using second-order statistics *IEEE Trans. Signal Process.* **45** 434–44

Berka C, Levendowski D J, Lumicao M N, Yau A, Davis G, Zivkovic V T, Olmstead R E, Tremoulet P D and Craven P L 2007 EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks *Aviation, Space, and Environmental Medicine* **78** B231–44

Blankertz B, Tomioka R, Lemm S, Kawanabe M and Robert-Müller K 2008 Optimizing spatial filters for robust EEG single-trial analysis *IEEE Signal Proc. Magazine* **25** 581–607

Boksem M A S, Meijman T F and Lorist M M 2005 Effects of mental fatigue on attention: an ERP study *Cogn. Brain Res.* **25** 107–16

Boonstra T W, Powell T Y, Mehrkanoon S and Breakspear M 2013 Effects of mnemonic load on cortical activity during visual working memory: linking ongoing brain activity with evoked responses *Int. J. Psychophysiology* **89** 409–18

Brouwer A-M, Hogervorst M A, van Erp J B F, Heffelaar T, Zimmerman P H and Oostenveld R 2012 Estimating workload using EEG spectral power and ERPs in the n-back task *J. Neur. Eng.* **9** 045008

Dijksterhuis C, de Waard D, Brookhuis K, Mulder B and de Jong R 2013 Classifying visuomotor workload in a driving simulator using subject specific spatial brain patterns *Frontiers in Neuroscience* **7** 149

Dussault C, Jouanin J, Philippe M and Guezennec C 2005 EEG and ECG changes during simulator operation reflect mental workload and vigilance *Aviation, Space, and Environmental Medicine* **76** 344–51

Fairclough S H 2008 Fundamentals of physiological computing *Interact. Comput.* **21** 133–45

Fu S and Parasuraman R 2007 Event-related potentials (ERPs) in neuroergonomics *Neuroergonomics: The Brain at Work* ed R Parasuraman and M Rizzo (New York, NY: Oxford University Press Inc.) pp 15–31

Gale A, Davies R and Smallbone A 1977 EEG correlates of signal rate, time in task and individual differences in reaction time during a five-stage sustained attention task *Ergonomics* **20** 363–76

George L and Lécuyer A 2010 An overview of research on passive brain-computer interfaces for implicit human-computer interaction *Proc. Int. Conf. Appl. Bionics. Biomech.*

George L, Marchal M, Glondu L and Lécuyer A 2012 Combining brain-computer interfaces and haptics: detecting mental workload to adapt haptic assistance *Haptics : Perception, Devices, Mobility, and Communication* ed D Hutchison *et al* vol 7282 (Berlin, Heidelberg: Springer Berlin Heidelberg) pp 124–35

Gevins A and Smith M E 2000 Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style *Cerebral Cortex* **10** 829–39

Gevins A and Smith M E 2007 Electroencephalography (EEG) in neuroergnomics *Neuroergonomics: The Brain at Work* ed R Parasuraman and M Rizzo (New York, NY: Oxford University Press Inc.) pp 15–31

Gomarus H K, Althaus M, Wijers A A and Minderaa R B 2006 The effects of memory load and stimulus relevance on the EEG during a visual selective memory search task: an ERP and ERD/ERS study *Clin. Neurophysiol.* **117** 871–84

Grimes D, Tan D S, Hudson S E, Shenoy P and Rao R P 2008 Feasibility and pragmatics of classifying working memory load with an electroencephalograph *Proc. 26th Annu. SIGCHI Conf. on Human Factors in Computing Systems, ACM* pp 835–44

Heger D, Putze F and Schultz T 2010 Online workload recognition from EEG data during cognitive tests and human-machine interaction *Lecture Notes in Artificial Intelligence: 6359, KI 2010: Advances in Artificial Intelligence,* eds T Dillmann, J Beyerer, U Hanebeck and T Schultz (Heidelberg: Springer) pp 410–7

Hoffmann U, Vesin J and Ebrahimi T 2006 Spatial filters for the classification of event-related potentials *European Symp. on Artificial Neural Networks* pp 47–52

Holm A, Lukander K, Korpela J, Sallinen M and Müller K M I 2009 Estimating brain load from the EEG *ScientificWorld J.* **9** 639–51

Hotelling H 1936 Relations between two sets of variates *Biometrika* **28** 321–77

Kaida K, Takahashi M, Åkerstedt T, Nakata A, Otsuka Y, Haratani T and Fukasawa K 2006 Validation of the Karolinska sleepiness scale against performance and EEG variables *Clin. Neurophysiol.* **117** 1574–81

Kok A 2001 On the utility of P3 amplitude as a measure of processing capacity *Psychophysiol.* **38** 557–77

Koles Z and Flor-Henry P 1981 Mental activity and the EEG: task and workload related effects *Med. Biol. Eng. Comput.* **19** 185–94

Miller M W, Rietschel J C, McDonald C G and Hatfield B D 2011 A novel approach to the physiological measurement of mental workload *Int. J. Psychophysiol.* **80** 75–8

Missonnier P, Deiber M-P, Gold G, Millet P, Gex-Fabry Pun M, Fazio-Costa L, Giannakopoulos P and Ibañez V 2006 Frontal theta event-related synchronization: comparison of directed attention and working memory load effects *J. Neur. Transm.* **113** 1477–86

Mühl C, Jeunet C and Lotte F 2014 EEG-based workload estimation across affective contexts *Frontiers in Neurosciences* **8** 114

Natani K and Gomer F E 1981 Electrocortical activity and operator workload: a comparison of changes in the electroencephalogram and in event-related potentials *McDonnell Douglas Technical Report* (Long Beach CA: McDonnell Douglas Corporation)

Nicolas-Alonso L F and Gomez-Gil J 2012 Brain computer interfaces, a review *Sensors* **12** 1211–79

Ossandón T, Jerbi K, Vidal J R, Bayle D J, Henaff M-A, Jung J, Minotti L, Bertrand O, Kahane P and Lachaux J-P 2011 Transient suppression of broadband gamma power in the default-mode network is correlated with task complexity and subject performance *J. Neurosci.* **31** 14521–30

Parasuraman R, Christensen J and Grafton S 2012 Neuroergonomics: the brain in action and at work *NeuroImage* **59** 1–3

Pearson K 1901 On lines and planes of closest fit to systems of points in space *Phil. Mag.* **2** 559–72

Rivet B, Souloumiac A, Attina V and Gibert G 2009 xDAWN algorithm to enhance evoked potentials: application to brain-computer interface *IEEE Trans. Biomed. Eng.* **56** 2035–43

Roy R N, Bonnet S, Charbonnier S and Campagne A 2013 Mental fatigue and working memory load estimation: interaction and implications for EEG-based passive BCI *Proc. IEEE Eng. Med. Biol. Soc. Conf.* pp 6607–10

Roy R N, Bonnet S, Charbonnier S and Campagne A 2015a Enhancing single-trial mental workload estimation through xDAWN spatial filtering *Proc. IEEE Conf. Neur. Eng.* pp 360–3

Roy R N, Bonnet S, Charbonnier S, Jallon P and Campagne A 2015b A comparison of ERP spatial filtering methods for optimal mental workload estimation *Proc. IEEE Conf. Eng. Med. Biol. Soc.* 7254–57

Schober F, Schellenberg R and Dimpfel W 1995 Reflection of mental exercise in the dynamic quantitative topographical EEG *Neuropsychobiology* **31** 98–112

Schultheis H and Jameson A 2004 Assessing cognitive load in adaptive hypermedia systems: physiological and behavioral methods *Lect. Notes Comput. Sci.* **3137** 225–34

Schultze-Kraft M, Gugler M, Curio G and Blankertz B 2012 Towards an online detection of workload in industrial work *Proc. IEEE Conf. Eng. Med. Biol. Soc.*

Spüler M, Walter A, Rosenstiel W and Bogdan M 2014 Spatial filtering based on canonical correlation analysis for classification of evoked or event-related potentials in EEG data *IEEE Trans. Neural. Syst. Rehabil. Eng.* **22** 1097–103

Sternberg S 1966 High-speed scanning in human memory *Science* **153** 652–4

Stipacek A, Grabner R, Neuper C, Fink A and Neubauer A 2003 Sensitivity of human EEG alpha band desynchronization to different working memory components and increasing levels of memory load *Neurosci. Lett.* **353** 193–6

Ullsperger P, Freude G and Erdmann U 2001 Auditory probe sensitivity to mental workload changes - an event-related potential study *Int. J. Psychophysiol.* **40** 201–9

van Erp J, Lotte F and Tangermann M 2012 Brain-computer interfaces: beyond medical applications *Computer* **45** 26–34

Young M S, Brookhuis K A, Wickens C D and Hancock P A 2015 State of science: mental workload in ergonomics *Ergonomics* **58** 1–17

Zander T O and Kothe C 2011 Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general *J. Neural. Eng.* **8** 025005